11. The Lunar Sample Preliminary Examination Team, "Preliminary Examination of Lunar Samples from Apollo 12," Science, 167, 1325 (1970).

12. D. H. Smith, "Mass Spectrometry," Chapter X in Guide to Modern Methods of Instrumental Analysis, T. M. Gouw, Ed., Wiley-Interscience, New York, 1972.

13. D. H. Smith, R. W. Olsen, F. C. Walls and A. L. Burlingame, "Real-time Mass Spectrometry: LOGOS--A Generalized Mass Spectrometry Computer System for High and Low Resolution, GC/MS and Closed-Loop Applications," Anal. Chem., 43, 1796 (1971).

14. A. L. Burlingame, J. S. Hauser, B. R. Simoneit, D. H. Smith, K. Biemann, N. Mancuso, R. Murphy, D. A. Flory and M. A. Reynolds, "Preliminary Organic Analysis of the Apollo 12 Cores," Proceedings of the Apollo 12 Lunar Science Conference, E. Levinson, Ed., M.I.T. Press, Cambridge, Mass., 1971, p. 1891.

15. D. H. Smith, "A Compound Classifier Based on Computer Analysis of Low Resolution Mass Spectral Data," Anal. Chem., 44, 536 (1972).

16. D. H. Smith and G. Eglinton, "Compound Classification by Computer Treatment of Low Resolution Mass Spectra-Application to Geochemical and Environmental Problems," Nature, 235, 325 (1972).

17. D. H. Smith, N. A. B. Gray, C. T. Pillinger, B. J. Kimble and G. Eglinton, "Complex Mixture Analysis - Geochemical and Environmental Applications of a Compound Classifier Based on Computer Analysis of Low Resolution Mass Spectra," Adv. in Org. Geochem., 1971, p. 249.

18. D. H. Smith, B. G. Buchanan, R. S. Engelmore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference, VIII. An Approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids," J. Amer. Chem. Soc., 94, 5962 (1972).

19. D. H. Smith, A. M. Duffield and C. Djerassi, "Mass Spectrometry in Structural and Stereochemical Problems, CCXXII. Delineation of Competing Fragmentation Pathways of Complex Molecules from a Study of Metastable Ion Transitions of Deuterated Derivatives," Org. Mass. Spectrom., 7, 367 (1973).

20. P. Longevialle, D. H. Smith, H. M. Fales, R. J. Highet and A. L. Burlingame, "High Resolution Mass Spectrometry in Molecular Structure Studies, V. The Fragmentation of Amaryllis Alkaloids in the Crinine Series," Org. Mass Spectrom., 7, 401 (1973).

21. B. R. Simoneit, D. H. Smith, G. Eglinton and A. L. Burlingame. "Applications of Real-time Mass Spectrometric Techniques to Environmental Organic Geochemistry, II. San Francisco Bay Area Waters," Arch. Env. Contam and Tox., 1, 193 (1973).

22. D. H. Smith, B. G. Buchanan, R. S. Engelmore, H. Adlercreutz and C. Djerassi, "Applications of Artificial Intelligence for Chemical inference, IX. Analysis of Mixtures Without Prior Separation as Illustrated for Estrogens," J. Amer. Chem. Soc., 95, 6078 (1973).

23. D. H. Smith, B. G. Buchanan, W. C. White, E. A. Feigenbaum, J. Lederberg and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference, X. INTSUM - A Data Interpretation and Summary Program Applied to the Collected Mass Spectra of Estrogenic Steroids," Tetrahedron, 29, 3117 (1973).

24. G. Loew, M. Chadwick and D. H. Smith, "Applications of Molecular Orbital Theory to the Interpretation of Mass Spectra. Prediction of Primary Fragmentation Sites in Organic Molecules," Org. Mass Spectrom., 7, 1241 (1973).

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME Sridharan, Natesa S.	TITLE Research Associate	BIRTHDATE *(Mo., Day, Yr.)* 10/2/46
PLACE OF BIRTH *(City, State, Country)* Madras, India	PRESENT NATIONALITY *(If non-U.S. citizen, indicate kind of visa and expiration date)* India; 5/73-U.S. permanent residence	SEX ☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Indian Institute of Technology, Madras India	Bachelor of Technology	1967	Electrical Engineering
State University of New York, Stony Brook	M.S.	1969	Computer Science
	Ph.D.	1971	Computer Science

HONORS

University Fellow - 1968-1971, SUNY Stony Brook; Graduate Assistant - 1967-1968, SUNY Stony Brook; Siemens' Award (awarded for top rank in Electrical Engineering) - 1967, ITT Madras; National Merit Scholarship - 1963-1967, ITT Madras

MAJOR RESEARCH INTEREST Computer Application in Chemistry and Medicine	ROLE IN PROPOSED PROJECT Research Associate

RESEARCH SUPPORT *(See instructions)*

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

1971-present	Research Associate, Heuristic Programming Project, Stanford University
1970-1971	Consultant, IAC Computer Corp., Long Island, N.Y.

Sridharan, N.S., "An Application of Artificial Intelligence to Organic Chemical Synthesis" Doctoral Thesis, State University of New York at StonyBrook, 1971.

Sridharan, N.S., "Search Strategies of Organic Chemical Synthesis", Third International Joint Conference on Artificial Intelligence (3IJCAI), Stanford, 1973

Sridharan, N.S. (co-author), "Heuristic DENDRAL: Analysis of Molecular Structure", Proc. NATO Advanced Study Institute, Amsterdam, 1973.

Sridharan, N.S. (co-author), "Heuristic Theory Formation", Machine Intelligence, Volume 7, Edinburgh, 1972.

U. S. GOVERNMENT PRINTING OFFICE : 1971 O - [illegible]

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME Brown, Harold D.	TITLE Associate Professor	BIRTHDATE *(Mo., Day, Yr.)* July 12,1934
PLACE OF BIRTH *(City, State, Country)* South Bend, Indiana	PRESENT NATIONALITY *(If non-U.S. citizen, indicate kind of visa and expiration date)* U.S.	SEX ☒ Male □ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
University of Notre Dame, Notre Dame, Indiana	M.Sc.	1963	Mathematics
Ohio State University, Columbus, Ohio	Ph.D.	1966	Mathamatics
(No Baccalaureate Degree)			

HONORS

Summa Cum Laude - Notre Dame

MAJOR RESEARCH INTEREST Applied Discrete Mathematics - Computer Science	ROLE IN PROPOSED PROJECT Research Associate

RESEARCH SUPPORT *(See instructions)*

Principal Investigator, NSF-GP-16793 (Expires March, 1974)

Pending Proposal NSF (Proposed starting date September, 1974)

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

Visiting Associate Professor, Computer Science, Stanford University , 1971-72, 1973-present
Associate Professor, Mathematics, Ohio State University, 1966-
Visiting Professor, Mathematics, Rhine,Westf. Tech. Hoch., Aachen, 1972 and 1973
Visiting Member, Courant Institute, New York University, 1967-68
Instructor/Assistant Professor, Assistant Chairman, Mathematics, Ohio State U., 1963-66
Assistant to the Chairman, Mathematics, University of Notre Dame, 1960-63
Director or Associate Director, NSF-SSTP, 1964-70

Publications

Near Algebras, Ill. J. Math. 12(1968), Pg. 215.

Distributor Theory in Near Algebras, Comm. Pure App. Math. XXI(1968), Pg. 535.

An Algorithm for the Determination of Space Groups, Math. Comp. 23(1969), Pg. 499.

Some Empirical Observations on Primitive Roots, with H. Zassenhaus, J. Number Theory 3(1971), Pg. 306.

A Generalization of Farey Sequences, with K. Mahler, J. Number Theory 3(1971), pg.364.

Basic Computations for Orders, Stanford CS Memo STAN-CS-72-208.

An Application of Zassenhaus' Unit Theorem, Acta Arith. XX(1972), Pg. 154.

Integral Groups I: The Reducible case, with J. Neubüser and H. Zassenhaus, Numer. Math. 19(1972), Pg. 386.

Integral Groups II: The Irreducible Case, with J. Neubüser and H. Zassenhaus, Numer. Math. 20(1972), Pg. 22.

Integral Groups III: Normalizers, with J. Neubüser and H. Zassenhaus, Math. Comp. 27(1973), Pg. 167.

Constructive Graph Labeling Via Double Cosets, with L. Hjelmeland and L. Masinter, Discrete Math. in press and Stanford CS Memo STAN-CS-72-318.

An Algorithm for the Construction of the Graphs of Organic Molecules, with L. Masinter, Discrete Math. in press and Stanford CS Memo STAN-CS-73-261..

The Crystallographic Groups of 4-dimensional Space, with J. Neubüser, H. Wondratschek and H. Zassenhaus, Wiley-Interscience in press.

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME DROMEY, Robert Geoffrey	TITLE Research Associate	BIRTHDATE *(Mo., Day, Yr.)* 11/21/46
PLACE OF BIRTH *(City, State, Country)* Castlemaine, Victoria, Australia	PRESENT NATIONALITY *(If non-U.S. citizen, indicate kind of visa and expiration date)* Australian, J-1 Visa, Exp. 10/8/74	SEX ☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Swinburne College of Technology, Melbourne, Australia	Diploma of Appl. Chem.	1968	Chemistry
La Trobe University Melbourne, Australia	Ph.D.	1973	Molecular Science

HONORS CSIRO Postdoctoral Studentship
Commonwealth Postgraduate Research Scholarship
Walter Lindrum Memorial Scholarship
Equivalent of First Class Honors Master of Science Preliminary (1969)

MAJOR RESEARCH INTEREST Application of Artificial Intelligence Techniques to Bio-Medical and Chemical Problems.	ROLE IN PROPOSED PROJECT Research Associate

RESEARCH SUPPORT *(See instructions)*

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

1973 DENDRAL Project, Stanford University, Computer Science Department
1973 Software Development for Graphics Systems, LaTrobe University, Computer Centre
1969-73 Construction, development and applications of an on-line photoelectron spectrometer LaTrobe University, Chemistry Department
1969-73 Application of Deconvolution Techniques to the Processing of Experimental Data.

Publications:

"Deconvolution and Its Application to the Processing of Experimental Data", Intl. Journal of Mass Spectrometry and Ion Physics, 1970, 4. (co-author).

"Inverse Convolution in Mass Spectrometry", Intl. Jnl. Mass Spec. Ion Phys., 1971, 6. (co-author).

"A Combined Time Averaging-Deconvolution Technique Applied to Electron Impact Ionization Efficiency Curves", Internation Journal of Mass Spectrometry & Ion Physics, 1971, 6. (co-author).

"The Perfect Direction and Velocity Focus at 254°34' in a Cylindrical Electrostatic Field", Reviews of Scientific Instruments, 1973, 44. (co-author).

"Detection of Spin-Orbit Splitting in the Photoelectron Spectrum of O_2^+ by Deconvolution", Chem. Physics Letters (in press), 1973. (co-author)

"The Effect of Finite Line Widths on the Interpretation of Photoelectron Spectra", Journal of Electron Spectroscopic (accepted for publication). (co-author).

"An On-line Ultraviolet Photoelectron Spectrometer for High-Resolution Studies of Molecular Structure", Australian Journal of Chemistry (accepted for publication). (co-author).

"Photoelectron Spectroscopic Correlation of the Molecular Orbitals of the Alkanes and Alkyliodides", Journal of Molecular Structure (submitted for publication). (coauthor).

"Comparison of the Photoelectron Spectra and the Photoionization Efficiency Curves for the Alkyliodides", Transactions of the Faraday Society (submitted for publication). (co-author).

"A Convolution-Deconvolution Algorithm Using Fast Fourier Transforms), Decuscope, 1973 (in press).

RESEARCH PLAN

BIOMOLECULAR CHARACTERIZATION: ARTIFICIAL INTELLIGENCE
A Program of Resource-Related Research

I. INTRODUCTION

This renewal application is intended to sustain and augment the capabilities of the mass spectrometry (MS) program which has served as a major institutional resource at Stanford for some years. With previous support from NASA and NSF it has made possible a highly interdisciplinary set of research projects ranging over: artificial intelligence (AI) in biomolecular characterization, natural product chemistry, clinical biochemical studies on steroids, and the mechanisms of molecular fragment formation in mass spectrometry. While the facility equipment for mass spectrometry has been funded mostly by other agencies, connected research programs embrace several NIH research projects as well. In addition, this activity was closely coupled with the ACME Medical School computer resource (1966-1973) and will have similar associations with the new AIM-SUMEX computer resource recently finded by the BRB (see Section I.C).

Previous support reflects the diversified facets of this interdisciplinary research. NASA has supported projects in new instrumentation, including the initial mass spectrometer-computer link, NSF has supported chemical research, and ARPA has supported our artificial intelligence research and initial application to mass spectrometry. Overall cutbacks have forced NASA to reduce funding for this area of research despite their interest. Under ARPA support to Drs. Feigenbaum and Lederberg for AI research, the DENDRAL program became recognized as one of the most successful AI applications programs. However, ARPA is chartered to fund frontier computer science research and no longer provides funds for the DENDRAL applications programs. ARPA has indicated a reluctance to continue funding to this group for the theory formation work in chemistry, although we expect to continue to receive ARPA support for more theoretical aspects of our research program (e.g., automatic programming).

We previously submitted a comprehensive proposal to the NIH (RR-00785, 3/28/73) which included an application for the AIM-SUMEX computing resource and a renewal of the existing DENDRAL grant (RR-00612). This proposal was approved for 5 years by the National Advisory Research Resources Council. Certain reservations were, however, communicated to us: they concerned especially what we must agree was an ambitious effort to close the control loop for "intelligent automation" whose costs overreached the immediate utility of the expected result. During subsequent discussions with the Biotechnology Resources Branch, taking into account the council review and a number of diverse policy issues, we agreed administratively to segment the two components of the original proposal. The AIM-SUMEX portion of the original proposal (excluding DENDRAL) was recently funded for 5 years as a national resource for artificial intelligence in medicine. The present proposal for resource-related research in biomolecular characterization and artificial intelligence is an elaboration of the DENDRAL portion incorporating intensive reexamination and revision of the previous proposal.

With the differentiation of priorities represented by AIM-SUMEX, the Genetics Research Center (GRC), and continuing work on artificial intelligence under Dr. Feigenbaum's leadership, the present renewal application places more emphasis than heretofore on real-world oriented applications. Correspondingly, we have agreed that it is now more appropriate that Dr. Djerassi should be designated as Principal Investigator in this phase of our work.

As outlined in section B.2, the interests and responsibilities of Professors Djerassi (Chemistry), Feigenbaum (Computer Science) and Lederberg (Genetics) have been closely interdigitated. With their further connections with many colleagues, these programs enjoy a high degree of university-wide participation. For example, the Genetics Department is also closely affiliated with Biology, Biochemistry, Pediatrics, Psychiatry and Medicine through joint appointments or joint research projects or both. This breadth would be difficult to obtain except at a few institutions where the medical school is both academically and geographically integrated with the university to the degree that characterizes the Stanford University environment.

GLOSSARY OF ABBREVIATIONS

ACME	- Advanced Computer for Medical Research (Nih-funded computer resource, 1968-1973)
AI	- artificial intelligence
AIM-SUMEX	- A comprehensive computer resource intended to serve the national requirement for artificial intelligence in medicine. This will be implemented at the Stanford University facility called AIM-SUMEX
ARPA	- Advanced Research Projects Agency of the Department of Defense.
BRB	- Biotechnology Resources Branch
13CMR	- carbon-13 magnetic resonance
GC	- gas chromatography or gas chromatograph
GRC	- Genetics Research Center (Stanford, J. Lederberg, Principal Investigator; NIGMS-approved and awaiting funding. Grant #P01-GM 20832-01)
HRMS	- high resolution mass spectrometry
IR	- infra-red
IRL	- Instrumentation Research Laboratory (Stanford Genetics Department)
LRMS	- low resolution mass spectrometry
MCD	- magnetic circular dichroism
MS	- Mass spectrometry or mass spectrometer
NASA	- National Aeronautics & Space Administration
NMR	- nuclear magnetic resonance
NSF	- National Science Foundation
ORD	- optical rotatory dispersion
PL/ACME	- a modified version of the PL-1 computer language (for the Stanford ACME computer facility)
SUMEX	- Stanford University Medical Experimental Computer Resource (NIH funded computer resource, 1973-1978)
UV	- ultra-violet

A. OBJECTIVES:

Core Research.
The funds now applied for would permit

1) the continued funding of the MS laboratory as a biomolecular characterization resource;

2) advancement of laboratory instrumentation capability in specific areas of GC-HRMS and the exploitation of metastable peak analysis.

3) the further development of AI computer techniques to match the instrumentation. This work will emphasize practical utilization for applications in biomolecular characterization connected with other on-going biomedical research programs. It will include, for example, a) the analysis of mixtures by GC/MS; b) metastable peak analysis for difficult problems of pure compounds and of mixtures not readily separable by GC; c) optimized data analysis for characterization of MS peaks and d) heuristic analysis of spectra for the molecular ion composition.

Our project is the only systematic effort, to our knowledge, currently underway in this country for computer assisted structure elucidation. Subsequent to our early publications, an intensive program has been mounted in Japan in similar areas. This situation may be contrasted with computer assisted organic synthesis, an area receiving considerable attention from several research groups. These capabilities can be beneficially provided to a wider community via the AIM-SUMEX resource. Research on the emulation of human intellect by computer programs will undoubtedly influence the efficiency with which chemical research can be applied to ever more complex problems of health, e.g., intermediary metabolism and its pathologies; environmental influences on health; the development and critical validation of new therapeutic agents.

The achievement of these objectives depends on the continued maintenance and development of the DENDRAL AI programming system (see below). The advent of the AIM-SUMEX facility will remove some of the serious computational limits on the exercise of this system that have delayed recent progress.

Education.
In our university setting, pre-doctoral and post-doctoral education of course constitutes a part of our mission. As far as is practically possible, research participation in the DENDRAL program has been coupled with dissertation work by graduate students and post-doctoral research experience respectively. Examples of people (and their research area) whose education has been enhanced in this way are the following:

Graduate Students: J. Simek, pedagogical aspects of the structure generator; Wai Lee Tan, synthesis of new estrogen compounds; H. Eggert, 13CMR of amines and steroidal ketones; C. Van Antwerp, 13CMR of steroidal alcohols; C. Farrell, theory formation from mass spectral data; L. Masinter, development of the structure generator; M. Stefik, AI applications to chemistry.

Postdoctoral Fellows: G. Dromey, theory formation from analytical data; R. Gritter, mass spectral fragmentation of biologically active steroids; R. Carhart, analysis of 13CMR spectra by DENDRAL-like programs; S. Hammerum, development of better fragmentation rules for progesterones.

Formal organization.
This project has been a long-term commitment of Djerassi, Lederberg and Feigenbaum functioning in effect as co-investigators. We coordinate our activities with day-to-day contacts in the pursuit of convergent research objectives. In the light of the extension of our collaborative activity during the last two years, we are now organizing a formal advisory group to include, in addition to ourselves, H. Cann, J. Barchas, and E. Van Tamelen. This group will advise the principal investigators on the direction of the program with respect to allocating available facilities and seeking out and helping other collaborators. This designation simply recognizes the fact that many of our colleagues have already been engaged in relevant collaborative research with us. A MS resource has recently been funded at the University of California/Berkeley, under the direction of Dr. A.L. Burlingame. Drs. Djerassi and Burlingame have recently engaged in some collaborative research which was made more successful by the sharing of facilities and expertise available at one institution but not at the other. We would hope to maintain and strengthen these contacts to avoid unnecessary duplication of effort.

We plan to discuss with Dr. A.L. Burlingame the most appropriate procedures for coordinating the related activities of our respective programs at the University of California/Berkeley and here. This may take the form of reciprocal membership in advisory committees.

The "hardware resource" to which this application is pegged has been identified as the MS facility. While these instruments alone represent an investment of over $300,000, funded previously by several agencies, they do not represent the most important resource. We would use this designation instead for the working team led by the princioal and co-investigators. The skills embraced by this group include, as mentioned, computer science, structural organic chemistry, molecular biology, instrumentation engineering and a wide range of other disciplines. They are represented not only in the principal professors but in a diversified and accomplished professional research staff (see Budget Justification). The program for which funds are now requested is the vital means by which the interests of this group can be sustained in a coordinated effort that would be very costly both in funds and in time if it had to be reconstructed from scratch. Without the financial support now requested, this line of collaborative research will have to be abandoned, with it a unique style of interdisciplinary collaboration, and the MS facility will be terminated.

B. BACKGROUND AND RATIONALE

1. The Structure Elucidation Problem

a) The General Problem. Analysis of molecular structure is a major activity in our program of resource related research. For the specific task of elucidating molecular structures, i.e., the topology of atom-to-atom connectivities, analysts utilize a mixture of information derived from chemical procedures and spectroscopic techniques. Each item of information, if not redundant or uninterpretable, contributes to the solution of the problem. Chemists draw upon a tremendous body of specific knowledge about the task area (e.g., clinical chemistry, biochemistry), molecular structure, spectroscopic techniques, etc., in order to piece together this information and infer the structure of molecules. These features, and the relative simplicity of the final concept of a structure, make the problem particularly well-suited for applications of the techniques of AI to assist research workers performing the task.

b) Djerassi's Laboratory. Professor Djerassi has been concerned with structure elucidation problems since the beginning of his chemical research. His activities at Stanford have been concerned heavily with the application of particular spectroscopic techniques to structural studies of biomedically important compounds. These techniques include optical rotatory dispersion (ORD) and, more recently, magnetic circular dichroism (MCD) (both of them supported initially by the NIH). Since 1961 he and his group have also been concerned with MS because of the power of the technique, in terms of specificity and sensitivity, as an analytical tool for structure elucidation. Four books and approximately 250 articles on MS have been published by him and his colleagues.

The technique of MS does not suffice for all structure determination problems, but it is a very powerful tool in areas where there exists a body of knowledge about the MS behavior of related molecules. When sample size is limited MS may well be the only technique that can be utilized. The recent availability of high resolution mass spectrometers has made HRMS the technique of choice for many applications because under ideal conditions the exact mass number uniquely specifies the the empirical formula of a molecule or fragment. On a parallel course, the technique of GC/MS, routinely available with low resolution mass spectrometers (GC/LRMS), has revolutionized investigations wherever complex mixtures are encountered. All of the above considerations argue that an extension of MS at Stanford to provide routine GC/LRMS and GC/HRMS analyses would be the next logical step to assist researchers depending on this facility for solutions of their structure elucidation problems.

2. Historical Background

a) Mass Spectrometry Laboratory. Prior to the existing DENDRAL grant, the groundwork was laid for computerization of the existing mass spectrometers, an Associated Electrical Industries MS-9 high resolution mass spectrometer and an Atlas CH-4 low resolution mass spectrometer. This work, supported primarily by NASA via the

Instrumentation Research Laboratory (IRL) in the Department of Genetics, resulted in link-up to the then existing ACME computer facility via a PDP-11 mini-computer which acted as a buffer between the spectrometers and ACME. Initial data acquisition and reduction programs were written for the system and utilized on a limited basis. The funding of the DENDRAL proposal, NIH grant RR-612 (May 1,1971-present) in conjunction with additional resources provided by the IRL resulted in a major improvement to these capabilities. The fruits of these efforts are described under section I.B.3 (below).

b) Summary of Early DENDRAL Development.
In 1964, Lederberg devised a notational algorithm for chemical structures (termed DENDRAL) that allowed questions of molecular structure to be framed in precise graph-theoretic terms. (Refs. 1,3-5,12). He also showed how to use the DENDRAL algorithm to generate complete and irredundant lists of structural isomers. (Refs. 1,6).

In 1965-66 Lederberg and Feigenbaum began exploring the idea of using the isomer generator in an artificial intelligence program - searching the space of possible structures for plausible solutions to a problem much as a chess-playing program searches the space of legal moves for the best moves. (Refs. 7,12). This approach guarantees that every possible solution to a problem is considered - either implicitly, as when whole classes of unstable structures are rejected, or explicitly, as when complete molecules are tested for plausibility. In either case, an investigator easily determines the criteria for rejection and acceptance and knows that no possibilities have been forgotten. This approach also guarantees that structures appear in the list only once - that automorphic representations of the same complex molecule have not been included. In both these respects the computer program has an advantage over manual approaches to structure elucidation.

c) Initial collaboration with Djerassi. (Refs. 14,15,19, 20,21,22,24).
Lederberg and Feigenbaum realized that (a) only through application to real problems could the AI approach be materially advanced and critically evaluated, and (b) MS appeared to be a fruitful applications area. MS appeared to be an excellent problem area because of the close relationship between spectral fragmentation patterns and molecular structure for many classes of molecules. Djerassi's interest and expertise - and daily interaction between members of his group and the AI group - led to a series of joint publications describing the approach and initial results of the programs. The success of these collaborative efforts led to the proposal to the NIH for initial funding to extend these efforts.

d) Efforts Under NIH Funding for DENDRAL. (Refs. 25-41).
The initial funding by NIH provided the opportunity to upgrade the instrumentation and computer programs. In particular we were able to mount a concerted project on both the analysis of mass spectra of biomedically important compounds and the mathematical aspects of molecular structure. Progress reports to the NIH describe this research in detail. The most recent annual report appears in Appendix B. A series of publications directed to audiences both in computer science and chemistry are listed in the bibliography. The following section (Section 3) summarizes the capabilities for

structure elucidation which, in themselves, constitute an important result of past work.

e) Related Research.
An important side effect of the DENDRAL project is the extent to which additional research was inspired and carried out to fill gaps in existing knowledge. This research, not supported by the DENDRAL grant, has been beneficial to on-going DENDRAL work, and vice-versa. Publications which have arisen from this research are listed in the bibliography (Refs. 58-70). A brief review of these publications should indicate the need for precise specification of the knowledge elicited from chemists and used in computer programs. As an example, consider the description and application of an early algorithm for generation of cyclic structural isomers (21). This paper considered the problem of spectroscopic differentiation of isomers of $C_6H_{10}O$. Unsaturated ethers fall in one of the classes of isomeric compounds which must be considered, but the MS of unsaturated ethers had not been investigated systematically. This work was subsequently carried out in Professor Djerassi's laboratory independently of DENDRAL support, but of benefit to DENDRAL (62). Other examples will be found in the Bibliography (Refs. 58-70).

3. Existing Capabilities

We have worked to develop distinctive capabilities for molecular structure elucidation, bringing together a high quality HRMS system and AI programs applied to biomolecular characterization. The feasibility of our analytical approach has been demonstrated in several problem areas, based upon the development both of a MS system and a general set of computer programs for use in new areas.

The principal capabilities are summarized below. These are now in being and were developed primarily under NIH funding to this project, with additional support supplied by ARPA and NASA in specific areas. (These agencies have reduced funding levels for this work because overall cutbacks have forced NASA to cut out this area of research despite their interest and ARPA is chartered to provide funds for frontier computer science research but not for applications. Thus the NIH is the principal of support for future development of applications programs in the interdisciplinary area of artificial intelligence/health related chemical problems.)

a. HRMS System and Coupled GC/LRMS System.
We have coupled the NIH-supported Varian-MAT 711 High Resolution Mass Spectrometer with a Hewlett Packard Gas Chromatograph and demonstrated its utility for GC/LRMS analysis of such difficult analytic problems as the free sterols (i.e., not derivatized) isolated from marine and other sources. Advanced data reduction techniques for this instrument were written for use with the ACME computer system (360/50) and now exist in Stanford's new 370/158 which continues to support the PL/ACME language. GC/HRMS scans on extracts from urine and amniotic fluid demonstrated this system's capability to provide high quality mass measurements on complex mixtures obtained from biological sources. An example of one GC/HRMS run on the amino acid fraction of amniotic fluid is presented below (Sec. III.D).

b. DENDRAL Structure Generator (Refs. 1-6,14,31,37,38,40,41)
The DENDRAL Structure Generator program accomplishes exhaustive and irredundant generation of isomers, with and without rings. This program guarantees consideration of every candidate structure - either implicitly, as when whole classes of structures are forbidden, or explicitly, as when individual compounds in a class are specified. It corresponds to the "legal move generator" of computerized chess playing and other heuristic programs.

c. DENDRAL Planner (Refs. 25,28,33)
We have written a very general set of computer programs for determining structural features from analytical data in well-defined areas. Such general planning programs have been written for low and high resolution MS, interpreted proton NMR spectroscopy and 13CMR data.

d. INTSUM (Refs. 26,29,34,35)
INTSUM is a computer program that aids in finding interpretive rules for MS. The program interprets a large collection of MS data according to criteria specified by the investigator. Then it summarizes the data to show which of the possible interpretations seem most plausible.

e. RULEGEN (Refs. 26,35)
RULEGEN is the current rule generation program that suggests various rules of interpretation for the MS data summarized by INTSUM. Although not finished, the program can provide useful assistance in practical theory formation.

f. Ancillary Techniques
1. The MS facility provides other types of experiments in MS, including ultra-high resolution measurements (masses determined via peak matching), defocussed metastable ion determinations (Barber-Elliott technique) and low ionizing voltage experiments. These data are utilized by both scientists and programs where appropriate.
2. Additional computer programs provide added problem-solving assistance.
   a. Predictor program for predicting major features of mass spectra.
   b. Programs for drawing and displaying chemical structures.
   c. Subroutines developed in conjunction with or existing as parts of the Structure Generator for problems of partitioning, construction of vertex-graphs, and constructive graph labelling. These can be applied to answer certain questions of isomerism which do not require the complete generator. For example, the labelling algorithm can list all structures resulting from substituting sites of a carbocyclic skeleton with stated numbers of different functional groups.

g. Other Spectroscopic Techniques
Available to us are the facilities of Professor Djerassi's laboratory for work requiring additional spectroscopic data. Also available on a fee for service basis are extensive spectroscopic facilities (NMR, I.R., and U.V.) of the chemistry department. These would be utilized for collecting additional data on particular structure problems and gathering data on known compounds (particularly in the area of 13CMR) as the AI programs become knowledgable about other spectroscopic information.

h. Chemical Facilities

The staff and facilities of the chemistry department represent substantial synthesis capabilities and general chemical know-how. This resource can be called upon to provide assistance in synthesis of model or labelled compounds, derivatization of mixtures, and so forth. For example, a graduate student in chemistry is presently engaged in thesis research dealing with the laboratory synthesis of a new estrogen metabolite strongly suspected to be a component of certain pregnancy urines. The previously proposed structure of this compound was one of the candidate structures inferred by the planner in a study of estrogen mixtures (11-dehydroestradiol-17-alpha, ref. 33).

4. User Community

Economic utilization of existing and proposed facilities can be realized by sharing them with a community of users. Lacking supplementary funds that would be needed for a comprehensive, major service facility, this community will include the following groups, but will be informally available to others.

A. Stanford Community

i) Stanford Chemistry Department (except for Hodgson, all are heavily supported by the NIH in their research efforts) Letters of interest are attached to the proposal in Appendix A.

Prof. C. Djerassi - Steroids, marine sterols
Prof. W. Johnson - steroids
Prof. E. Van Tamelen - steroids, triterpenoids, other natural products
Prof. H. Mosher - natural products (e.g., marine toxins)
Prof. K. Hodgson - biological ligands, ligand-metal complexes
Prof. J. Collman - cytochrome P450 models

ii) Stanford Medical School Collaborators

The following research projects in the Stanford Biomedical Community will furnish samples for mass spectrometric analysis under the present proposal. Attached to this proposal (Appendix A) are copies of the letters of interest in the proposed facility received from the principal investigators of these grants.

Dr. James R. Trudell, Department of Anesthesia, Stanford University School of Medicine. Drug metabolite identification in humans.

Dr. Irene S. Forrest, Biomedical Research Laboratory, Veterans Administration Hospital, Palo Alto. Drug metabolite identification in humans.

Dr. I. Rabinowitz and D.I. Wilkinson, Department of Dermatology, Stanford University School of Medicine. Prostaglandins.

Prof. Eugene D. Robin, Department of Respiratory Medicine, Stanford University School of Medicine, Ratio of NAD+/NADH in cells by measuring ratio of oxidized to reduced redox pairs.

Dr. Leo E. Hollister, Veterans Administration

Hospital/Department of Medicine, Stanford University School of Medicine. Metabolism of Marihuana.

Dr. Hiram H. Sera, Pharmacy Department, Stanford University Hospital. Drug Identification.

Dr. Sumner M. Kalman, Department of Pharmacology, Stanford University School of Medicine. Drug and drug metabolite identification.

Dr. Jack Barchas, Department of Psychiatry, Stanford University School of Medicine. Neurotransmitters and related compounds in man.

Dr. Keith A. Kvenvolden, Chemical Evolution Branch, NASA Ames Research Center, Mountain View, Calif. Amino acids, acids in geochemical samples, structure of products formed from electrical discharges in gas mixtures.

Dr. William R. Fair, Department of Urology, Stanford University School of Medicine. Identification of the prostatic antibacterial factor; polyamines (putrescine, spermine, spermidine) in body fluids of patients with prostatic carcinoma.

Besides the user projects just summarized, other major prospects are in sight. At the time of writing, the chair of pharmacology is vacant. Conversations with the leading candidate have indicated a deep-seated interest in GC/HRMS as the principal analytical tool for broad ranging studies of drug metabolism in man.

B. Extramural Users

The development of the techniques of ORD, MS and MCD at Stanford has been paralleled with extensive sharing of these resources nation- and world-wide in collaborative research efforts, without any additional funding. Rather than provide routine service, experience has shown that discretionary selection of problems results in better utilization of our people and instrumentation resources. We would extend this provision of services including available computer programs, to a limited number of extramural users. Note, for example, our successful collaboration with Professor Adlercreutz, Meilahti Hospital, University of Helsinki, on the identification of estrogens from body fluids utilizing the AI planning program (ref. 33).

C. Relationship to AIM-SUMEX and the Genetics Research Center

The present application is strengthened by two research projects related to, but not overlapping, the proposed research of this grant.

1) AIM-SUMEX (NIH RR-00785, Oct. 1, 1973, thru July 31, 1978, Principal Investigator, J. Lederberg). This is a resource grant to establish a national facility for applications of artificial intelligence in medicine (AIM). Our own use of this facility will include SUMEX PDP-10 computer time and file storage necessary to run the DENDRAL artificial intelligence programs. This support will be furnished without charge to the present proposal. It represents an annual investment of about $100,000 in computer time equivalent value.

The AIM-SUMEX computing facility is shared equally between a national user community (AIM) and a Stanford Medical School community. The DENDRAL research will be supported out of the Stanford portion. The AIM service will be administered under the policy control of a national advisory committee and will be implemented over a national computer network. AIM-SUMEX provides the means for members of the national user community interested in structure elucidation to access the DENDRAL programs.

2) Genetics Research Center (NIH PO1-GM 20832-01 - approved by the NIGMS Council, awaiting funding, Principal Investigator, J. Lederberg). This research proposal is a comprehensive grant which would support interdepartmental research at the Stanford Medical School in Medical Genetics, Pediatrics and other clinical applications. A section of that proposal concerns the use of GC/LRMS for screening body fluids for evidence of inborn errors of metabolism. (This project grew out of the initial DENDRAL grant, one of the research goals of which was the analysis of body fluids using GC/MS). This research on inborn metabolic errors will be conducted jointly in the Stanford Departments of Genetics and Pediatrics using existing equipment (Finnigan 1015 Quadrupole mass spectrometer, Varian Aerograph GC and a PDP-11/20 based data system).

We appreciated the value of GC/HRMS analyses of selected extracts of body fluids (i.e., those containing metabolites not identified by routine GC/LRMS data) when formulating the Genetics Research Center proposal. Accordingly, a small amount of funding was there requested for recording selected GC/HRMS data on the GC/Varian MAT 711 mass spectrometer in the Department of Chemistry. If these funds are awarded, we will negotiate with NIH a suitable elimination of this minor overlap with the present budget.

II. SPECIFIC AIMS

The specific aims enumerated in this section will be pursued in the highly inter-disciplinary manner that has characterized the DENDRAL project from the start of its NIH support. The aims are not disjoint,but interactive and inter-dependent. For example, the power of MS and, potentially, other spectroscopic techniques, can be enhanced by the use of computer programs to perform various aspects of structure elucidation and theory formation. From the standpoint of computer science, one measure of the utility of techniques of artificial intelligence is how well they perform in real-world applications. It is necessary in the development of these programs to have a source of data and informed, involved team-mates able to criticize methods and results. The aims are elaborated in the methods section.

We have attempted to keep the proposal to a readable length. Therefore, some detail has been omitted. However, many details can be found in the biliography and we are prepared to provide additional information during the site visit.

1. Enhance the power of the MS resource.

The existing MS resource, together with computer programs which exist or which are proposed (see Aim 2, below), is capable of solving some of the structure elucidation problems of the user community given computer support for data collection and reduction. We refer specifically to the areas of GC/LRMS and routine, batch HRMS samples. We believe that many of the problems of the user community require more powerful techniques (see Section III). These techniques, specifically GC/HRMS and semi-automatic metastable defocussing, can be provided with a minimum of cost and effort, thus enhancing considerably the capabilities of the resource.

Our first aim is to provide the resource with adequate computer support (replacing the previous ACME system) to enable collection and reduction of mass spectral data including low and high resolution scans and data on defocussed metastable ions.

We propose to develop this computer support in the ways described below. (these aims are written to include the work necessary to implement the extended PDP-11/20 computer system. A description of the rationale for this choice is provided in Section III.A and the specific augmentations in the Budget Justification).

A) Convert existing, proven data acquisition and reduction programs from the PL/ACME language into Fortran, consistent with time-critical assembly language programs for data acquisition and instrument control. These programs will be written in Fortran to enhance compatibility with the computer systems of other users of such packages.

B) Modify these programs, as required, to handle acquisition and reduction of frequent or repetitive HRMS scans with selected instrument performance feedback to the operator, and to take advantage of the expanded capabilities of the extended 11/20 system. Prototype GC/HRMS systems have been developed at Stanford and elsewhere, but this type of facility (in contrast to GC/LRMS) is not now available to the Stanford community. When this system is developed, service will be available to the Stanford community and research collaborators and, if our resources permit, to any scientist requesting assistance. In many instances this type of collaboration will require far more involvement of convergent interests, efforts and skills than merely running samples on request. We have in mind the chemical and eventually biological interpretation of the analytical data as a matter of joint concern, as appropriate.

We have previously illustrated the advantages of high resolution mass spectral data in the computer analysis of mass spectra (e.g., ref. 28). Also, we have previously shown that the same program can deal with analyses of mixtures without prior separation especially when additional data (e.g., from selected metastable defocussing experiments) were provided (Ref. 33). We wish to use the MS resource and the computer program in further studies of mixtures of compounds which are difficult or impossible to separate by GC. The advent of routine systems for high pressure liquid chromatography have made many of these separations possible, but the liquid chromatograph is not presently interfaced to the MS.

Many of the problems of the user community require analysis of complex mixtures which are amenable to treatment by GC/MS techniques. We feel that where sample quantities permit, acquisition of GC/HRMS data is highly desirable. These data can be provided by the resource supplemented with computer support (above).

We propose to continue tests of the GC/MS combination, operating under moderately high mass resolutions (5000-10000), to define in detail the optimum operating conditions of the GC/HRMS combination. This will provide the necessary information on maximum practical sensitivity to be expected. This information can then be used in collaboration with the user community for sample preparation.

The GC/HRMS system would normally be operated at reduced mass spectrometer resolutions to maximize sensitivity. We have existing multiplet resolution programs to increase the resolving power of the MS. We propose to provide the multiplet resolution program with heuristic guidance based on compositional variations inferred from molecular ions or other singlet peaks. For example, a resolving power of 10,000 is barely sufficient to resolve ions which differ by CH_2 vs. N (delta m = 0.012) for ions of about mass 100. Although it will resolve CH_4 vs. O doublets (delta m = 0.037) at this mass, it will not resolve closer doublets such as C_3N vs. H_2O_3 (delta m = 0.003). We can provide exhaustive tabulations of multiplets by mass separations (based on ref. 30) which can be used by the multiplet resolution program.

We have previously indicated the power of metastable ion information in the operation of our programs for structure elucidation (refs. 28, 33). We have extended one of our programs (the MS predictor program) to propose metastable defocussing experiments in order to avoid collection of unnecessary data (see Aim 2, below). Although we can collect these data (Barber-Elliott technique) manually on our existing Varian MAT-711 MS, this is an exceedingly wasteful operation, both in terms of sample consumption and time. We propose to implement some automation of collection of these data on metastable ions. We also propose to begin preliminary investigation of alternative modes of metastable ion determination (see Methods (Sec. III), below).

2. Develop performance and theory formation programs to assist in the solution of structure elucidation problems in biomedicine.

Computer programs have already been written for analysis of low and high resolution mass spectra, for generation of acyclic and cyclic molecular structures, for labelling structural skeletons with atoms, for analyzing 13CMR spectra of amines and for interpretation and summary of large volumes of data gathered on model compounds (see Existing Capabilities above, for references). We wish to increase the utility of these programs by providing interactive facilities that allow easier access to them, by increasing their generality and power, and by supplementing them with new reasoning programs.

Performance Programs:

The current structure generator program will be subjected to further detailed tests before using it for structure determination problems.

A new algorithm for generating cyclic skeletons (with no multiple bonds) will be programmed and checked. The algorithm is written and informally proved. A formal proof will be devised as well. This algorithm represents one very powerful approach to the problem of implementation of constraints, as discussed in the following paragraph.

The generating programs will be modified to allow isomer generation within constraints. Different kinds of constraints can be inferred from different kinds of spectroscopic data. We intend to give the program knowledge of a variety of these.

The Planner programs that infer constraints from mass spectrometry data will be broadened to include additional knowledge about the spectral behavior of classes of compounds of relevance to the NIH-sponsored research of the user community. In addition, we will add the capability for utilization of information about chemical isolation procedures (e.g., one expects acidic and neutral compounds in solvent extraction of acidified body fluids) and relative GC retention times (e.g., to admit the possibility of homologous series).

We propose to implement a more general method for inferring the identity of the molecular ion whether or not this appears explicitly in the spectrum. This information is important for the successful operation of the structure generator and the planner. We want the program to use whatever information is available and not depend, as it currently does, on having knowledge of the structural class together with inference rules for that class.

Interface routines will be written to make it easier for other scientists to use these programs. We have to wait for an interactive system before starting this: AIM-SUMEX will be ideal. Input/output routines will be crucial to easy use of the system. However, we also want to give users the facility to understand the system's reasoning steps so they can take advantage of it.

In addition to making the computer programs available through AIM-SUMEX, we would like to translate parts of the LISP code into another language - for reasons of both efficiency and exportability. We have talked with computer professionals at IBM Research Center about using the APL language. FORTRAN, ALGOL and PL/1 are other languages whose merits for our purposes we will explore.

We wish to continue a low-level of effort on computer programs that interpret other kinds of spectroscopic data. Planning programs similar to the MS Planner could be written for automatic analysis of data from other spectroscopic techniques(e.g., IR, UV), as we have illustrated for 13CMR (ref. 39).

The structure generator's view of chemical structure is topological and is presently unconstrained by bond lengths and angles. Because stereochemical considerations are frequently important in structure elucidation, we propose to begin consideration of stereochemistry in the structure generation and evaluation processes.

A program with detailed knowledge about information obtainable from various spectroscopic techniques could be written to examine a list of candidate solutions and propose experiments necessary and sufficient to distinguish among them. The program would represent an extended Predictor (e.g., ref. 27). We have a first version of a program that suggests "crucial" metastable peaks to be sought in order to distinguish among candidate structures. Work on this program will continue at a low level of activity, possibly expanding into areas other than MS. One topic we will continue to pursue is our collaborative effort with Dr. Gilda Loew, Genetics Department, on the potential application of molecular orbital theory to prediction of mass spectra (ref. 71).

Theory Formation Programs:

The rule formation program (RULEGEN) will be extended so that it can search a larger space of rules. Present a priori constraints on the rule generation give us a search reduction from tens of millions to a thousand possible rules. Even though search heuristics now allow efficient search of these possibilities, we want to be able to deal with much larger spaces efficiently, as when the number of primitive predicates is drastically increased.

The RULEGEN program will be modified so that complex fragmentation and rearrangement processes are manipulated nearly as easily as simple fragmentations. The program currently finds fragmentation rules involving one or two bonds, possibly followed by hydrogen migration. In the case of cyclic systems such as estrogens, however, the program must be able to work with sets of three or more bonds in some cleavages.

Interactive programs will be provided on AIM-SUMEX for the investigator to query the rule generation program. For example, many questions now arise about the program steps by which the program infers the rules it suggests as explanations of the regularities. Why, for example, was some particular rule not considered plausible?

New data will have to be selected in order to test the rules and to differentiate among competing rules. We will write a program that suggests new experiments (i.e., new data to obtain), depending on the nature of the existing rules.

The test phase of the theory formation program will be written as an evaluation function of each rule against new data. Insofar as any new experiments are "crucial" experiments, the evaluation function may merely reject a proposed rule. Mostly, however, rules will have to be evaluated against new data along many dimensions: frequency, strength of evidence, uniqueness, simplicity, and the like.

We wish to experiment with the whole theory formation program to determine the critical aspects of our design. For example, (1) how sensitive is the program to discrepancies, inconsistencies and errors in the data? (2) how well can the program find rules within a slightly different model of chemistry? (3) how well can the program perform with one pass through the data, or several passes? and (4) how critical are the principles of theory formation?

3. Apply the structure elucidation techniques - both instrumentation and computer programs - to biomedically relevant compounds.

Our own interests are in elucidating the structures of, and understanding the MS of, marine sterols, hormonal steroids, and compounds isolated from human body fluids that can be associated with genetic disorders (from research in the GRC). In addition, we will be working closely with members of the Stanford Medical School and Chemistry Department - in particular those mentioned above (Section I.B.4) - on their structure elucidation problems in which MS will be used. Although most users expect to require HRMS and GC/HRMS data, some of their problems will be attacked utilizing GC/LRMS techniques and library search through (usually) restricted libraries of mass spectral data. We propose to investigate some extensions to the technique of library search (see Methods) to complement our existing and planned DENDRAL programs. We plan to continue our exchange of mass spectral data and library search information as we have previously done with Dr. S. Markey (University of Colorado Medical School) and Dr. F. W. McLafferty (Cornell University).

As in the past, attention to new biomedical research problems will lead to increased capabilities in the computer programs. We require close communication with the people engaged in the research so that the programs actually assist the researcher while increasing in power. Collaborative proposals have come out of such past DENDRAL sponsored work, for example, large portions of the GRC proposal and a proposal for 13CMR research.

We envision the interaction and collaboration with the user community to involve the following: